# Speech Processing 2010/11

3rd Test

May 26th 2011

Please identify this form with your name and student number in the reserved space at the bottom. The answers to multiple-choice questions will only be accepted if inserted in the appropriate place. Wrong answers to these questions will be penalized. The phonetic symbols should use the SAMPA alphabet (Lisbon accent).

1. Classify as True (T) or False (F)

   (a) In earlier speech recognition systems, endpoint detection typically involved energy and duration.

   (b) Cepstral mean subtraction is typically done to achieve robustness against environmental characteristics as well as microphone characteristics.

   (c) Low frequency words often deviate from their canonic pronunciation.

   (d) Witholding a part of the training data in order to estimate interpolation weights may be done for smoothing both acoustic and language models.

   (e) Covariance matrices used in Gaussian mixture models may be diagonal in order to restrict the number of parameters to train.

   (f) Cross word modeling is easier to do with context-independent models than with context-dependent models.

   (g) Retraining models for new speakers may be done by using the segments that have been recognized with high confidence using speaker-independent models.

   (h) In evaluating the probability of an observation sequence of T frames being produced by an HMM with N states, the forward model involves in the order of NT operations.

   (i) An HMM left-to-right topology without skips involves training more parameters than one with skips.

   (j) LVCSR systems typically show the same WER for real-time and ten-times-real-time performance.

2. The *de facto* standard features used in speech recognition: (mark as True (T) all that apply)

   (a) model spectral info

   (b) are typically extracted every 30 ms

   (c) model temporal info

   (d) represent the log-energy output of a filter-bank with uniform channel spacing

   (e) model cepstral info

   (f) are generally uncorrelated

3. What is the typical perplexity range for a trigram language model of 20k vocabulary size in romance languages (choose one)?

   (a) PP $< 40$

   (b) $40 <$ PP $< 150$

   (c) $150 <$ PP $< 200$

   (d) PP $> 200$

4. Write the expression that relates the perplexity with the entropy of a language model.

5. Consider an HMM discrete model with 2 states, which models the extraction of colored balls from two urns. In the first urn, there are 10 different balls, all equally probable, one of them is black. In the second urn, there are 20 balls of different colors, 10 of them are black, all equally probable. The probabilities of transition between states are:

   - Prob(staying in the first urn)=0.6
   - Prob(staying in the second urn)=0.7
   - Prob(moving from 1st to 2nd urn)=0.4
   - Prob(moving from 2nd to 1st urn)=0.3

   The first extraction is made from the first urn. All balls are put back in the corresponding urns after each extraction. What is the probability of getting a black ball in the second extraction? (Only indicate the computations.)

6. When speaking in loud noise, there is an involuntary tendency of speakers to increase the intensity of their voice, to enhance its audibility. This change includes not only loudness but also other acoustic features such as pitch and rate and duration of sound syllables.

   (a) What is the name of this effect?

   (b) Does automatic speech recognition improve in this type of speech, relative to normal speaking conditions? (Y/N)

7. Name:

   (a) two vector quantization methods

   (b) two smoothing techniques

   (c) two enhancement techniques

   (d) two distance measures between typical features used in earlier recognition systems

8. The following extract was produced by a speaker independent large vocabulary continuous speech recognition system for a broadcast news piece:

   *Para as eleições geral ainda não terminou neste momento parece a*
   *todo o minuto a qualquer momento, que o príncipe William e Kate menos tanto, entra*
   *nada dia de Westminster para ensaiar ao detalhe aquilo que vai acontecer*
   *na própria sexta-feira.*

   The corresponding manual transcription is below:

   *Olha José o ensaio geral ainda não terminou. Neste momento espera-se a*
   *todo o minuto, a qualquer momento, que o príncipe William e Kate Middleton entrem*
   *na Abadia de Westminster para ensaiar ao detalhe aquilo que vai acontecer*
   *na própria sexta-feira.*

   Ignoring punctuation and capitalization, compute the corresponding values of H (”correct”), D (”deletions”), S (”substitutions”), I (”insertions”), N (”total”), %Corr, %Acc and %WER.

9. How many phones differ in the hypothesis and the reference in the first two words of the third line? Justify, indicating the respective broad phonetic transcriptions.

10. Consider the training corpus that consists of the following sentences:
    *Harry is younger than Fred.*
    *Fred is older than Ron.*
    *Harry and Ron are older than Ginny.*

*Ron and Fred have red hair.*
*Ginny and Percy have red hair.*

Consider the test sentence:
*Percy and Ron are older than Fred.*

(a) Compute the number of unigrams, bigrams and trigrams of the training corpus (excluding $< /s >< s >$ transitions), and the dimension of the vocabulary.

(b) Compute the probability of the test utterance using a bigram language model without any type of smoothing.

(c) Compute the probability of the test utterance using a bigram language model with add-one smoothing.

(d) What is the maximum number of words in a sentence with non-zero probability according to a bigram model without smoothing?

(e) What are the most frequent trigrams in the training corpus?

This page is intentionally left empty.

**Test 3 - Answers**

| | |
|---|---|
| Name: | |
| Number: | |

1. (3.0 val.) Indicate T or F:

| a | b | c | d | e | f | g | h | i | j |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | |

2. (1.2 val.) Mark as T all that apply:

| a | b | c | d | e | f |
|---|---|---|---|---|---|
| | | | | | |

3 (1.0 val.)

| |
|---|
| |

4 (1.0 val.)

| |
|---|
| |

5. (2.0 val.)

| |
|---|
| |

6. (1.2 val.)

| | |
|---|---|
| a | |
| b | |

7. (2.4 val.)

| | |
|---|---|
| a | |
| b | |
| c | |
| d | |

8. (3.0 val.)

| H | D | S | I | N | % Corr | % Acc | % WER |
|---|---|---|---|---|---|---|---|
| | | | | | | | |

9. (1.2 val.)

| | |
|---|---|
| a | |
| b | |

10. (4.0 val.)

| | |
|---|---|
| a) | |
| b) | |
| c) | |
| d) | |
| e) | |